

# Truncated Maximum Likelihood, Goodness of Fit Tests and Tail Analysis

Christian Gouriéroux\*

Joanna Jasiak†

This version: March 11, 1998

(First draft: September 28, 1997)

---

\*CREST and CEPREMAP, *e-mail*: [gouriero@ensae.fr](mailto:gouriero@ensae.fr).

†York University, *e-mail*: [jasiakj@yorku.ca](mailto:jasiakj@yorku.ca)

**Abstract**

Truncated Maximum Likelihood,  
Goodness of Fit Tests and Tail Analysis.

We propose a new method of tail analysis for data featuring a high degree of leptokurtosis. Heavy tails can typically be found in financial series, like for example, the stock returns or durations between arrivals of trades. In our framework, the shape of tails can be assessed by fitting some selected pseudo-models to extremely valued observations in the sample. A global examination of the density tails is performed using a sequence of truncated pseudo-maximum likelihood estimators, called the tail parameter function (t.p.f.). In practice, data often exhibit local irregularities in the tail behaviour. To detect and approximate these patterns we introduce the local parameter function (l.p.f.), a method of tail analysis involving a selected interval of extreme observations. An immediate extension of the pseudo-value based approach to density analysis yields a new procedure for testing the goodness of fit. We also develop a new nonparametric estimator of the density function. The method is applied to unequally spaced high frequency data. We study an intradaily series of returns on the Alcatel stock, one among the most frequently traded stocks on the Paris Stock Exchange.

Keywords: Goodness of fit test, tail analysis, truncated maximum likelihood, kernel, Hill estimator.

## Résumé

Maximum de vraisemblance tronqué,  
tests d'adéquation et analyse de queues.

Nous proposons une nouvelle méthode d'analyse des queues de distribution. Un effet leptokurtique important est généralement mis en évidence sur la plupart des séries financières, qu'il s'agisse de séries de rendements ou de durées entre échanges. Dans notre approche la forme des queues est étudiée en ajustant divers pseudo-modèles aux données extrêmes de l'échantillon. Un examen global des densités de queues peut être effectué en utilisant une suite d'estimateurs du pseudo-maximum de vraisemblance tronqué, appelée fonction de paramètre de queue [t.p.f.]. Une approche alternative repose sur des analyses locales de la queue [l.p.f.]. Ces analyses permettent aussi d'introduire des tests d'ajustement et un nouvel estimateur à noyau de la densité. L'approche est appliquée à une série de rendements du titre Alcatel, l'un de plus échangés à la bourse de Paris.

Mots clés: Tests d'adéquation, analyse de queues, maximum de vraisemblance tronqué, noyau, estimateur de Hill.

# 1 Introduction

Since the seminal work of Mandelbrot (1963) on cotton prices, the presence of heavy tails in the marginal distributions of financial returns has been confirmed by a large number of empirical studies. This phenomenon has inspired a number of researchers seeking to provide a plausible explanation. Mandelbrot advanced a hypothesis of an underlying stable non-normal distribution of returns in a random walk model. Clark (1973) and Harris (1989) suggested that returns are generated by a mixture of normal distributions, with the rate of new information arrival acting as a stochastic mixing variable. The heavy tails may also be viewed as a consequence of conditional heteroscedasticity, due to time varying volatility of returns [Engle (1982)]. In the limiting case of strong persistence and presence of a unit root in the volatility equation, the marginal variance may even not exist [Nelson (1990)]. The presence of heavy tails was also documented in the conditional distributions of returns for some financial assets. This feature is crucial for risk management and should be accounted for in the risk control rules adopted by financial institutions. Indeed, standard rules [such as the Value at Risk] are implicitly based on a mean-variance approach, and are inappropriate when data frequently admit extreme values. The leptokurtosis of conditional distributions may also explain some other characteristics of the series of financial returns such as the long range of temporal dependence [Resnick, Samorodnitsky (1997)].

In general, we say that a probability distribution features a heavy tail if asymptotically it behaves like the Pareto distribution  $P[Y > y] = y^{-\gamma}L(y)$ ,  $\gamma > 0$ ,  $y > 1$ , where  $L$  is slowly varying:  $\lim_{t \rightarrow \infty} \frac{L(ty)}{L(t)} = 1$ . The statistical methods for tail analysis focus on the estimation of the tail index  $\gamma$ . In the case of i.i.d. observations, the estimators are generally computed from the order statistics:  $Y_{(1)} \geq \dots \geq Y_{(n)}$ . Two well-known estimators of  $\gamma$  are the Hill estimator [Hill (1975)]:

$$\frac{1}{\hat{\gamma}} = \frac{1}{k} \sum_{i=1}^k \log \left( \frac{Y_{(i)}}{Y_{(k+1)}} \right),$$

and the de Haan's moment estimator [de Haan (1970)]:

$$\frac{1}{\hat{\gamma}} = \frac{1}{k} \sum_{i=1}^k \left[ \log \left( \frac{Y_{(i)}}{Y_{(k+1)}} \right) \right]^r.$$

In this paper we propose a new method of tail analysis. We introduce a parametric family of distributions  $\mathcal{F} = \{f(y; \theta), \theta \in \Theta\}$  and determine the value of the  $\theta$  parameter such that  $f(y; \theta)$  is at the minimum distance from the empirical distribution of the observed  $y$ 's, for values sufficiently large  $y \geq c$ , or for  $y$  lying in a neighbourhood of large  $c$ . The tail behavior of the distribution is revealed by the behavior of  $\theta$  with respect to  $c$ .

The formulas of estimators and test statistics, and their properties are derived under the assumption of i.i.d. observations  $Y_1, \dots, Y_n$ . We present the asymptotic results under general regularity conditions.

In section 2 we introduce a partition of the real line into non overlapping intervals and allow the parameter  $\theta$  to vary independently between the intervals. The comparison of the full sample maximum likelihood estimators of  $\theta$  with the maximum likelihood estimators truncated over the intervals yields various goodness of fit tests. In section 3 we consider the pseudo maximum likelihood (henceforth P.M.L.) estimation of the  $\theta$  parameter from the data truncated by  $c$ . The limits of the P.M.L. estimators define a function of  $c$ , called the tail parameter function (t.p.f.). We show that the tail parameter function is constant only for a distribution in the  $\mathcal{F}$  family. We discuss conditions under which the t.p.f. indicates the true underlying distribution. In section 4, we introduce an estimator of the t.p.f., derive its asymptotic distributional properties and discuss the links with the Hill estimator. In section 5 we develop a local pseudo-maximum likelihood method for estimation of density functions. The basic idea consists in smoothing the estimator of  $\theta$  censored over an interval using a kernel. Next, the log-derivative of the unknown density function at a given point is approximated by the log-derivative of the pseudo density evaluated at the local estimator of  $\theta$ . The empirical results are discussed in section 6, where various methods proposed in the paper are applied to a series of financial returns. Some technical issues are explained in Appendices 1, 2 and 3. Section 7 concludes the paper.

## 2 Truncated Maximum Likelihood and Goodness of Fit Tests

### 2.1 Decomposition of the Log-Likelihood Function

Let us consider a family  $\mathcal{F}$  of distributions on  $\mathcal{Y} \in R$ , parametrized by  $\theta \in \Theta \subset R^p$ . The distributions in  $\mathcal{F}$  have positive p.d.f.  $f(\cdot; \theta)$  with respect to a dominating measure  $\mu$ . We denote by  $F(\cdot; \theta)$  and  $S(\cdot; \theta)$  the corresponding cumulative function and survivor function, respectively.

For a random variable  $Y$  with the distribution function  $f(\cdot; \theta)$ , and  $A_1, \dots, A_K$  a partition of  $\mathcal{Y}$ , we introduce the qualitative variable  $Z$  indicating the element of the partition containing  $Y$ :

$$Z = (Z_1, \dots, Z_K)' = [1_{A_1}(Y), \dots, 1_{A_K}(Y)]', \quad (2.1)$$

where  $1_A(Y) = 1$ , if  $Y \in A$  and  $1_A(Y) = 0$ , otherwise.

The marginal distribution of the vector  $Z$  is multinomial  $M(1; F(A_1; \theta), \dots, F(A_K; \theta))$ , with the probability function  $P[Z = z] = h(z; \theta)$ . We denote by  $f_k(y; \theta)$  the conditional density function of  $Y$  given  $Z_k = 1$ , (i.e.  $Y \in A_k$ ).

By applying the Bayes theorem we obtain the following decomposition of the conditional density function of  $Y$ :

$$\log f(y; \theta) = \log h(z; \theta) + \sum_{k=1}^K z_k \log f_k(y; \theta), \quad (2.2)$$

which implies:

$$-\frac{\partial^2 \log f(y; \theta)}{\partial \theta \partial \theta'} = -\frac{\partial^2 \log h(z; \theta)}{\partial \theta \partial \theta'} - \sum_{k=1}^K z_k \frac{\partial^2 \log f_k(y; \theta)}{\partial \theta \partial \theta'}. \quad (2.3)$$

The expectations of all terms in (2.3) with respect to the distribution  $f(\cdot; \theta)$  yield the following decomposition of the information matrix:

$$J_Y(\theta) = J_Z(\theta) + \sum_{k=1}^K J_{Y|k}(\theta), \quad (2.4)$$

where

$$\begin{aligned} J_Y(\theta) &= E_\theta \left[ -\frac{\partial^2 \log f(Y; \theta)}{\partial \theta \partial \theta'} \right], \\ J_Z(\theta) &= E_\theta \left[ -\frac{\partial^2 \log h(Z; \theta)}{\partial \theta \partial \theta'} \right], \\ J_{Y|k}(\theta) &= E_\theta \left[ -Z_k \frac{\partial^2 \log f_k(Y; \theta)}{\partial \theta \partial \theta'} \right]. \end{aligned}$$

Let us also introduce the matrix:

$$J_{Y|Z}(\theta) = J_Y(\theta) - J_Z(\theta) = \sum_{k=1}^K J_{Y|k}(\theta), \quad (2.5)$$

which is the information matrix based on the conditional distribution of  $Y$  given  $Z$ .

## 2.2 Extended Model

The model  $\mathcal{F}$  may be nested in a larger model  $\mathcal{F}^*$ , allowing for an independent variation of the parameters appearing in the marginal distribution of  $Z$  and in the conditional distributions of  $Z$  given  $Y \in A_k$ . We now consider the model  $\mathcal{F}^*$ , where

- the marginal distribution of  $Z$  is  $h(z; \gamma)$ ,
- the conditional distribution of  $Y$  given  $Y \in A_k$  is  $f_k(y; \theta_k)$ ,  $k = 1, \dots, K$ ,
- $\gamma, \theta_1, \dots, \theta_K$  are unconstrained with values in  $\Theta \subset R^p$ .

The parameters of this extended model can consistently be estimated from a sample  $Y_1, \dots, Y_n$  of i.i.d. variables by the maximum likelihood. The log-likelihood function to be maximized:

$$L_Y(\gamma; \theta_1, \dots, \theta_K) = \sum_{i=1}^n \log h(z_i; \gamma) + \sum_{k=1}^K \left( \sum_{i=1}^n z_{k,i} \log f_k(y_i; \theta_k) \right) \quad (2.6)$$

consists of two additive terms involving separately the parameters  $\gamma$  and  $\theta$ . This allows for distinct optimizations with respect to these parameters and ensures an asymptotic independence of the estimators of  $\gamma$  and  $\theta_1, \dots, \theta_K$ .

**Property 2.1:**

Let  $Y_1, \dots, Y_n$  be i.i.d. variables with a distribution in  $\mathcal{F}^*$ . Under standard regularity conditions, the unconstrained maximum likelihood estimators  $\tilde{\gamma}, \tilde{\theta}_1, \dots, \tilde{\theta}_K$  are consistent, asymptotically normal, and asymptotically independent:

$$\begin{aligned} \sqrt{n}[\tilde{\gamma} - \gamma] &\xrightarrow{d} N[0, J_Z(\gamma)^{-1}], \\ \sqrt{n}[\tilde{\theta}_k - \theta_k] &\xrightarrow{d} N[0, J_{Y|k}(\theta_k)^{-1}], \quad k = 1, \dots, K. \end{aligned}$$

## 2.3 Goodness of Fit Tests

In this subsection we present a procedure involving various pseudo-true value based tests to evaluate the fit of the model. Let us introduce the hypotheses:  $H = \{\gamma, \theta_1, \dots, \theta_K, \text{unconstrained}\}$ ,  $H_0^* = \{\theta_1 = \dots = \theta_K\}$ ,  $H_0 = \{\theta_1 = \dots = \theta_K = \gamma\}$ .

Under the hypothesis  $H_0$ , the common parameter value of  $\theta$  and  $\gamma$  can be estimated by the maximum likelihood based on the entire sample of  $n$  observations:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \log f(y_i; \theta). \quad (2.7)$$

Under  $H_0$ , this estimator is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N[0, J_Y(\theta)^{-1}]. \quad (2.8)$$

Under the hypothesis  $H_0^*$ , two estimation methods can be considered. One consists again on maximizing the likelihood function and yields the estimator of  $\theta^* = \theta_k$ ,  $k = 1, \dots, K$ . Alternatively, we can use the equivalent approximation:

$$\tilde{\theta}^* = \left[ \sum_{k=1}^K J_{Y|k}(\tilde{\theta}_k) \right]^{-1} \sum_{k=1}^K J_{Y|k}(\tilde{\theta}_k) \tilde{\theta}_k. \quad (2.9)$$

Under the hypothesis  $H_0$ ,  $\tilde{\gamma}$  and  $\tilde{\theta}^*$  are consistent estimators of  $\theta$ , asymptotically normal and independent, with the limiting variance matrices:

$$V_{as}[\sqrt{n}(\tilde{\gamma} - \theta)] = J_Z(\theta)^{-1}, \quad (2.10)$$

$$V_{as}[\sqrt{n}(\tilde{\theta}^* - \theta)] = J_{Y|Z}(\theta)^{-1}. \quad (2.11)$$

Let us now consider the sequence of nested hypotheses:

$$H_0 \subset H_0^* \subset H = \{f \in \mathcal{F}^*\},$$

and introduce test statistics  $\xi_{H_0|H}$ ,  $\xi_{H_0|H_0^*}$ ,  $\xi_{H_0^*|H}$  for testing  $H_0$  against  $H$ ,  $H_0$  against  $H_0^*$ ,  $H_0^*$  against  $H$  respectively.

The test statistics can be formulated as the following likelihood ratios:

$$\begin{aligned} \xi_{H_0|H} &= 2[L_Y(\hat{\theta}) - L_Y(\tilde{\gamma}; \tilde{\theta}_1, \dots, \tilde{\theta}_K)], \\ \xi_{H_0|H_0^*} &= 2[L_Z(\hat{\theta}) - L_Z(\tilde{\gamma})], \\ \xi_{H_0^*|H} &= 2[L_{Y|Z}(\hat{\theta}) - L_{Y|Z}(\tilde{\theta}_1, \dots, \tilde{\theta}_K)], \end{aligned}$$

where  $L_Y, L_Z, L_{Y|Z}$  are the log-likelihood functions corresponding to the observations of  $Y, Z$  and to the conditional model of  $Y$  given  $Z$  respectively. Standard results on likelihood ratio tests [Gouriéroux, Monfort (1997)] imply the following property:

**Property 2.2:**

Under the null hypothesis  $H_0$ , the statistics  $\xi_{H_0|H_0^*}$  and  $\xi_{H_0^*|H}$  are asymptotically independent, with asymptotic distributions:  $\chi^2(p)$  and  $\chi^2((K-1)p)$ , respectively.

The statistic  $\xi_{H_0|H} = \xi_{H_0|H_0^*} + \xi_{H_0^*|H}$  follows asymptotically a  $\chi^2(Kp)$  distribution.

It is possible to replace these test statistics by their asymptotic equivalents which are easier to compute. Such equivalent statistics are:



$$\xi_{H_0|H_0^*} = (\hat{\gamma} - \tilde{\theta}^*)' [\hat{V}(\hat{\gamma}) + \hat{V}(\tilde{\theta}^*)]^{-1} (\hat{\gamma} - \tilde{\theta}^*), \quad (2.12)$$

$$\xi_{H_0^*|H} = \sum_{k=1}^K (\tilde{\theta}_k - \tilde{\theta}^*)' (\hat{V}(\tilde{\theta}_k))^{-1} (\tilde{\theta}_k - \tilde{\theta}^*). \quad (2.13)$$

The extended model  $\mathcal{F}^*$  may itself be nested in a more general model  $\mathcal{F}^{**}$ , where the marginal distribution of  $Z$  is multinomial with parameters  $p_1, \dots, p_K$  constrained only by  $p_k \geq 0, \forall k, \sum_{k=1}^K p_k = 1$ , and where the conditional distributions of  $Y$  given  $Z$  are as in  $\mathcal{F}^*$ . The likelihood ratio statistic  $\xi^*$  for testing  $H = \{f \in \mathcal{F}^*\}$  against  $H^* = \{f \in \mathcal{F}^{**}\}$  is asymptotically equivalent to the standard chi-square statistic of the goodness of fit test. Under the null hypothesis  $H_0$ , the statistic  $\xi^*$  is asymptotically independent of  $\xi_{H_0|H_0^*}, \xi_{H_0^*|H}, \xi_{H_0|H}$  and follows the asymptotic  $\chi^2(K - p - 1)$  distribution. In some sense the statistics  $\xi_{H_0|H_0^*}, \xi_{H_0^*|H}$  are complementary to the standard chi-square goodness of fit test.

### 3 The Tail Parameter Function

In section 2 we showed the pseudo-true value based test statistics to assess the fit of the model. We now extend this approach and introduce a new method of tail analysis which instead of the pseudo-true values based on the entire sample involves tail based pseudo-true values.

#### 3.1 Definition

Let us consider a parametric family of density functions:  $\mathcal{F} = \{f(\cdot; \theta), \theta \in \Theta \subset R^p\}$ , with respect to the Lebesgue measure on  $\mathcal{Y} = R^+ = (0, +\infty[$ . We assume that the density functions in this family are positive.

For any probability distribution, with density function  $f_0(\cdot)$ , we can write the optimization problem:

$$\max_{\theta \in \Theta} E_0 \left[ 1_{Y \geq c} \log \frac{f(Y; \theta)}{S(c; \theta)} \right], \quad (3.1)$$

where  $c \in R^+$  and  $E_0$  is the expectation with respect to  $f_0 \cdot \mu$ . We assume:

**Assumption A.1.** The optimization problem (3.1) has a unique solution  $\theta(c; f_0; \mathcal{F})$  for any  $c \geq 0$ .

The parameter value  $\theta(c; f_0; \mathcal{F})$  is such that the parametrized distribution  $f(\cdot; \theta)$  provides the best approximation of the tail of the true distribution defined by the level  $c$ .

**Definition 3.1** Let us consider a distribution  $f_0$  satisfying A.1. The tail parameter function associated with  $f_0$  is the mapping  $\theta(\cdot; f_0; \mathcal{F})$  from  $R^+$  on  $\Theta$ .

For notational convenience we further denote the tail parameter function (t.p.f.) by  $\theta(\cdot; f_0)$ , omitting  $\mathcal{F}$ . Under standard regularity conditions,  $\theta(c; f_0)$  is the limit of a pseudo maximum likelihood estimator based on the pseudo model  $\mathcal{F}$  and a sample  $Y_1, \dots, Y_n$  with distribution  $f_0$ , and left truncated by  $c$ :

$$\theta(c; f_0) = \lim_{n \rightarrow \infty} \hat{\theta}_n(c), \quad (3.2)$$

where:

$$\hat{\theta}_n(c) = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \{1_{y_i \geq c} \log \frac{f(y_i; \theta)}{S(c; \theta)}\}. \quad (3.3)$$

### 3.2 Characterization of a Distribution of the Pseudo Model

In this subsection we show that the t.p.f. is constant only for distributions in  $\mathcal{F}$ .

#### Property 3.2:

Let us assume that the density functions are first order differentiable with respect to  $\theta$  and  $y$ , and that the parameter  $\theta$  is identifiable. The density  $f_0$  belongs to the family  $\mathcal{F}$  if and only if the tail parameter function is constant over  $c$ .

#### Proof:

##### Necessary Condition

If  $f_0$  belongs to  $\mathcal{F}$ , there exists a unique parameter value  $\theta_0$  such that  $f_0(\cdot) = f(\cdot; \theta_0)$ , and the t.p.f. is  $\theta(c; f_0) = \theta_0, \forall c$ .

##### Sufficient Condition

Let  $\theta^*$  the common solution of the optimization problems:

$$\begin{aligned} \theta^* &= \arg \max_{\theta} E_0[1_{Y > c} \log \frac{f(Y; \theta)}{S(c; \theta)}] \\ &= \arg \max_{\theta} [\int_c^{\infty} \log f(y; \theta) f_0(y) dy - S_0(c) \log S(c; \theta)]. \end{aligned}$$

The first order conditions are:

$$\int_c^{\infty} \frac{\partial \log f(y; \theta^*)}{\partial \theta} f_0(y) dy - S_0(c) \frac{\partial \log S(c; \theta^*)}{\partial \theta} = 0, \forall c.$$

By differentiating again with respect to the truncation level  $c$  we get:

$$-\frac{\partial \log f(c; \theta^*)}{\partial \theta} f_0(c) - \frac{dS_0(c)}{dc} \frac{\partial \log S(c; \theta^*)}{\partial \theta} - S_0(c) \frac{\partial^2 \log S(c; \theta^*)}{\partial c \partial \theta} = 0, \forall c.$$

$$f_0(c) \left[ -\frac{\partial \log f(c; \theta^*)}{\partial \theta} + \frac{\partial \log S(c; \theta^*)}{\partial \theta} \right] - S_0(c) \frac{\partial^2 \log S(c; \theta^*)}{\partial c \partial \theta} = 0, \forall c.$$

$$\frac{f_0(c)}{S_0(c)} = \frac{\partial^2 \log S(c; \theta^*)}{\partial c \partial \theta_j} \left[ -\frac{\partial \log f(c; \theta^*)}{\partial \theta_j} + \frac{\partial \log S(c; \theta^*)}{\partial \theta_j} \right]^{-1}, \quad \forall c, j = 1, \dots, p.$$

The hazard function corresponding to  $f_0$  is uniquely defined by  $f(c; \theta^*)$  and since the density function  $f(c; \theta^*)$  also satisfies the last equation, we conclude that  $f_0(\cdot) = f(\cdot; \theta^*)$ .

Q.E.D.

### 3.3 Identifying a Distribution by its Tail Parameter Function

In this subsection we focus on two important questions:

- (i) Does there exist pseudo-families  $\mathcal{F}$  such that  $\theta(\cdot; f_0)$  characterizes any  $f_0$  ?
- (ii) For a given pseudo-family  $\mathcal{F}$ , what condition satisfy the distributions identifiable by their t.p.f. ?

We give below some preliminary answers to these questions.

**Property 3.3:**

If  $Y$  is a positive variable and the pseudo-family is the family of exponential distributions:  $\mathcal{F} = \{f(y; \theta) = \theta \exp(-\theta y), \theta \in R^+\}$ , then

- (i) the t.p.f. is the expected residual life :  $\theta(c; f_0) = E_0((Y - c)|Y > c)$ ;
- (ii) the t.p.f. characterizes the distribution.

**Proof:**

- (i) The optimization problem

$$\max_{\theta} \{E_0[1_{Y>c} \log \theta] - E_0[1_{Y>c} \theta(Y - c)]\},$$

involves the first order condition:

$$\frac{1}{\theta} E_0(1_{Y>c}) - E_0[1_{Y>c} (Y - c)] = 0,$$

and yields the solution:

$$\theta(c; f_0) = E_0((Y - c)|Y > c).$$

- (ii) Conversely we get:

$$\begin{aligned} \theta(c; f_0) &= \frac{E_0[1_{Y>c} (Y - c)]}{E_0(1_{Y>c})} \\ &= \frac{\int_c^\infty (y - c) f_0(y) dy}{S_0(c)} \\ &= \frac{\int_c^\infty S_0(y) dy}{S_0(c)}. \end{aligned}$$

Therefore  $\theta(c; f_0)^{-1} = -\frac{d}{dc} [\log \int_c^\infty S_0(y)dy]$ , and, when  $\theta(\cdot; f_0)$  is known, the survivor function can be found by integration.

Q.E.D.

**Property 3.4:**

There exists a one to one relationship between the distribution  $f_0$  and its t.p.f. in a neighbourhood of the parametric pseudo-model  $\mathcal{F}$ .

**Proof:** The first order condition defining  $\theta(c; f_0)$  is:

$$\forall c: \int_c^\infty f_0(y) \frac{\partial}{\partial \theta} \log f[y; \theta(c, f_0)] dy - S_0(c) \frac{\partial \log S}{\partial \theta}[c; \theta(c, f_0)] = 0.$$

Let us consider a distribution  $f_0$  close to  $f(\cdot; \theta^*)$  and introduce  $\delta(c; f_0) = \theta(c, f_0) - \theta[c, f(\cdot; \theta_0^*)] = \theta(c, f_0) - \theta_0^*$ .

A first order expansion yields:

$$\begin{aligned} \forall c: \int_c^\infty [f_0(y) - f(y; \theta_0^*)] \frac{\partial}{\partial \theta} \log f(y; \theta_0^*) dy - [S_0(c) - S(c; \theta_0^*)] \frac{\partial}{\partial \theta} \log S(c; \theta_0^*) \\ + \delta(c; f_0) \left\{ \int_c^\infty f(y; \theta_0^*) \frac{\partial^2}{\partial \theta^2} \log f(y; \theta_0^*) dy - S_0(c) \frac{\partial^2}{\partial \theta^2} \log S(c; \theta_0^*) \right\} = 0. \end{aligned}$$

Consider now a given discrepancy function  $\delta(\cdot; f_0)$ . By differentiating the previous relation with respect to  $c$  we obtain a differential equation of the type:

$$[f_0(c) - f(c; \theta_0^*)]A(c) - [S_0(c) - S(c; \theta_0^*)]B(c) + C(c) = 0,$$

where  $A, B, C$  are unknown functions. This is a first order linear differential equation for the function  $S_0(\cdot) - S(\cdot; \theta_0^*)$ . The set of solutions is affine of dimension 1, and the integration constant is uniquely determined by the initial condition:  $S_0(-\infty) - S(-\infty; \theta_0^*) = 1 - 1 = 0$ .

Q.E.D.

The previous property is a local identifiability condition of the true density by its t.p.f.

### 3.4 The Form of the Tail Parameter Function in a Neighborhood of the Pseudo-Family

In this section we show how to evaluate the discrepancy function  $\delta(\cdot; f_0)$  introduced in the Property 3.4, in a neighborhood of the pseudo-family. The property 3.5 below follows directly from the expansion of the first order condition and corresponds to some standard properties of the maximum likelihood estimators under local alternatives.

**Property 3.5:**

Let us consider a parametric family:  $\tilde{\mathcal{F}} = \{f(\cdot; \theta, \beta), (\theta, \beta) \in \Theta \cdot B\}$ , and assume  $\mathcal{F} = \{f(\cdot; \theta) = f(\cdot; \theta, 0), \theta \in \Theta\}$ , and  $f_0 = f(\cdot; \theta_0^*, \beta)$  with small  $\beta$ . Then  $\theta(c; f_0) = \theta_0^* + \tilde{J}_{\theta\beta}(c)\beta$ , where  $\tilde{J}_{\theta\beta}(c)$  is the cross term of the information matrix corresponding to the model  $\mathcal{F}$  truncated at  $c$  and evaluated at  $(\theta_0^*, 0)$ .

## 4 Tail Analysis

In this section we present the t.p.f. method of tail analysis. We begin with the null hypothesis concerning the tail of the distribution:

$$\tilde{H}_0 = \{\exists c_0 : f_0(y) = f_0(y; \theta_0), \forall y \geq c_0\}.$$

This hypothesis may be tested by applying the procedure presented in section 2, based on a partition of the real line. Else, for an increasing sequence of  $c$  it can be verified if the consistent estimator of the t.p.f. remains approximately constant for large  $c$ . The second approach is similar to some classical methods of tail analysis. It provides exploratory plotting techniques, but should be used with caution. Its limitation results from a strong correlation of the estimated values of the t.p.f. at different points, due to the overlapping intervals.

### 4.1 Statistics Based on the Truncated Pseudo Maximum Likelihood

Let us consider a sample of i.i.d. variables  $Y_1, \dots, Y_n$  with an unknown distribution  $f_0$ . We are interested in testing the null hypothesis  $\tilde{H}_0$ . Since the t.p.f. is constant for  $c \geq c_0$ , it is natural to expect that a consistent estimator of  $\theta(\cdot; f_0)$  is constant as well. Such an estimator is:

$$\hat{\theta}_n(c) = \arg \max_{\theta} \sum_{i=1}^n [1_{y_i \geq c} \log f(y_i; \theta)] - \log S(c; \theta) \sum_{i=1}^n 1_{y_i \geq c}, \quad (4.1)$$

$c$  varying. This functional estimator is a stepwise function with jumps at each observed value. To see that, let us introduce the order statistics corresponding to the observations:

$$y_{(1)} \geq y_{(2)} \geq \dots \geq y_{(n)}.$$

The jumps arise at the values  $y_{(k+1)}$ ,  $k$  varying, and we get:

$$\hat{\theta}_{n,k} = \hat{\theta}_n(y_{(k+1)}) \quad (4.2)$$

$$= \arg \max_{\theta} \sum_{i=1}^k \log f(y_{(i)}; \theta) - \log S(y_{(k+1)}; \theta) k. \quad (4.3)$$

*Example 4.1:* When the pseudo-family  $\mathcal{F}$  is exponential family  $\mathcal{F} = \{f(y; \theta) = \theta \exp(-\theta y), \theta \in R^+\}$ , the estimator is:

$$\frac{1}{\hat{\theta}_{n,k}} = \frac{1}{k} \sum_{i=1}^k [y_{(i)} - y_{(k+1)}].$$

*Example 4.2 :* When the pseudo-family is Pareto family,  $\mathcal{F} = \{f(y; \theta) = \theta y^{-\theta-1}, \theta \in R^+\}$ , we get:

$$\frac{1}{\hat{\theta}_{n,k}} = \frac{1}{k} \sum_{i=1}^k \log \left[ \frac{y_{(i)}}{y_{(k+1)}} \right],$$

which is the Hill estimator [ Hill (1975)], with properties in the i.i.d. case discussed by Mason (1982).

## 4.2 Asymptotic Properties of the Estimated t.p.f. under the Null Hypothesis

The asymptotic behavior of  $\hat{\theta}_n(c_1), \dots, \hat{\theta}_n(c_J)$  for a finite set of truncation points:  $c_j, j = 1, \dots, J$  with  $c_j \geq c_0, \forall j$  follows from the standard results on the maximum likelihood estimators. We denote by  $J(c; \theta_0)$  the information matrix of the model truncated by  $c$ .

### Property 4.1:

Under the null hypothesis  $\tilde{H}_0$ :

(i) the estimator  $[\hat{\theta}_n(c_1), \dots, \hat{\theta}_n(c_J)]'$  converges to  $(\theta_0, \dots, \theta_0)$  and is asymptotically normal:  $\sqrt{n}[\hat{\theta}_n(c_1) - \theta_0, \dots, \hat{\theta}_n(c_J) - \theta_0]' \xrightarrow{d} N(0, \Omega)$ , where  $\Omega = A^{-1}BA^{-1}$ ,

$$A = \begin{bmatrix} J(c_1; \theta_0) & & 0 \\ & \ddots & \\ 0 & & J(c_J; \theta_0) \end{bmatrix},$$

$$B = [J(\sup(c_i, c_j); \theta_0)].$$

(ii) In particular, for a truncation point  $c$  we get:

$$\sqrt{n}[\hat{\theta}_n(c) - \theta_0] \xrightarrow{d} N[0, J(c; \theta_0)^{-1}].$$

## 5 Local P.M.L. Analysis of the Density Function

### 5.1 The Estimation Method

An alternative application of the approach based on approximating the unknown density function  $f_0$  by  $f(\cdot; \tilde{\theta}_k)$  on the interval  $A_k$  (see, section 2) is a non parametric estimation of the density

function  $f_0$ . Let us consider an interval  $A = [c - h, c + h]$ . The approximation of the  $\theta$  parameter is the solution of:

$$\begin{aligned}\tilde{\theta}_{c,h} &= \arg \max_{\theta} \left[ E_0 [1_{c-h < Y < c+h} \log f(Y; \theta)] - E_0 [1_{c-h < Y < c+h}] \log \int 1_{c-h < y < c+h} f(y; \theta) dy \right] \\ &= \arg \max_{\theta} E_0 \left[ \frac{1}{2h} 1_{c-h < Y < c+h} \log f(Y; \theta) \right] - E_0 \left[ \frac{1}{2h} 1_{c-h < Y < c+h} \right] \\ &\quad \log \int \frac{1}{2h} 1_{c-h < y < c+h} f(y; \theta) dy, \\ \tilde{\theta}_{c,h} &= \arg \max_{\theta} E_0 \left[ \frac{1}{h} K \left( \frac{Y - c}{h} \right) \log f(Y; \theta) \right] - E_0 \left[ \frac{1}{h} K \left( \frac{Y - c}{h} \right) \right] \log \int \frac{1}{h} K \left( \frac{y - c}{h} \right) f(y; \theta) dy, \\ &\hspace{15em} (5.1)\end{aligned}$$

where  $K(u) = \frac{1}{2} 1_{[-1,1]}(u)$ .

From the last formula, we derive a non parametric estimation method for the density function  $f_0$ . We consider a kernel satisfying the following assumptions:

*Assumptions concerning the kernel:*

**A.2**  $\int K(u) du = 1;$

**A.3**  $\int K(u) u du = 0;$

**A.4**  $\int K(u) u^2 du = \eta^2, \text{ (exists).}$

We propose to use the estimator defined by :

$$\tilde{\theta}_n(c) = \arg \max_{\theta} \left[ \sum_{i=1}^n \frac{1}{h} K \left( \frac{y_i - c}{h} \right) \log f(y_i; \theta) - \sum_{i=1}^n \frac{1}{h} K \left( \frac{y_i - c}{h} \right) \log \int \frac{1}{h} K \left( \frac{y - c}{h} \right) f(y; \theta) dy \right], \quad (5.2)$$

to estimate the p.d.f.  $f_0$  from the pseudo distribution after substituting  $\tilde{\theta}_n(c)$  for  $\theta$ . This approach is based on a local approximation of the p.d.f. by an element of the pseudo family  $\mathcal{F}$  and is an analogue of the locally weighted least square regression developed by Cleveland (1979) and Hardle (1990) in the context of non parametric estimation of the regression function [see also Gouriéroux, Monfort, Tenreiro (1994) for the extension to kernel M-estimators].

## 5.2 The Gaussian Pseudo-family

As an illustration of the method described above we examine a special case where the estimator  $\tilde{\theta}_n(c)$  admits an explicit form and can easily be interpreted. Let us consider a gaussian pseudo-family with the mean  $\theta$  :

$$f(y; \theta) = \phi(y - \theta),$$

where  $\phi$  is the p.d.f. of the standard normal distribution. Let us also use a gaussian kernel:  $K(\cdot) = \phi(\cdot)$ . We get:

$$\begin{aligned} \int \frac{1}{h} \phi\left(\frac{y-c}{h}\right) f(y; \theta) dy &= \int \frac{1}{h} \phi\left(\frac{c-y}{h}\right) \phi(y - \theta) dy \\ &= \frac{1}{\sqrt{1+h^2}} \phi\left(\frac{c-\theta}{\sqrt{1+h^2}}\right), \end{aligned}$$

and the estimator is:

$$\begin{aligned} \tilde{\theta}_n(c) &= \arg \max_{\theta} \sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i - c}{h}\right) \left[ -\frac{1}{2} \log 2\pi - \frac{(y_i - \theta)^2}{2} \right] \\ &\quad - \sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i - c}{h}\right) \left[ -\frac{1}{2} \log 2\pi - \frac{1}{2} \log(1+h^2) - \frac{(c-\theta)^2}{2(1+h^2)} \right]. \end{aligned}$$

It is equal to :

$$\tilde{\theta}_n(c) = \frac{1+h^2}{h^2} \frac{\sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right) y_i}{\sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right)} - \frac{1}{h^2} c.$$

We find that:

$$\tilde{\theta}_n(c) = \frac{1+h^2}{h^2} \tilde{m}_n(c) - \frac{1}{h^2} c,$$

where :

$$\tilde{m}_n(c) = \frac{\left(\sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right) y_i\right)}{\left(\sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right)\right)},$$

is the Nadaraya - Watson estimator of the regression function:  $m(c) = E[Y|Y = c] = c$ , [Nadaraya (1964), Watson (1964) ].

Moreover:

$$\begin{aligned} \tilde{\theta}_n(c) - c &= \frac{1+h^2}{h^2} (\tilde{m}_n(c) - c) \\ &= \frac{1+h^2}{h^2} \frac{\sum_{i=1}^n \frac{1}{h} (y_i - c) \phi\left(\frac{y_i-c}{h}\right)}{\sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right)} \\ &= (1+h^2) \frac{\frac{\partial}{\partial c} \left[ \sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right) \right]}{\left[ \sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i-c}{h}\right) \right]} \\ &= (1+h^2) \frac{\partial \log \tilde{f}_n(c)}{\partial c}, \end{aligned}$$



where  $\tilde{f}_n(c)$  is the (gaussian) kernel estimator of the density function  $f_0$ . Then, the asymptotic properties of  $\tilde{\theta}_n(c) - c$  are deduced directly from the asymptotic properties of  $\tilde{f}_n(c)$  and  $\frac{\partial \tilde{f}_n(c)}{\partial c}$  [Silverman (1986)]. In particular  $\tilde{\theta}_n(c)$  converges to  $c + \frac{\partial \log}{\partial c} f_0(c)$  when  $n$  tends to infinity and  $h$  tends to zero at appropriate rates.

### 5.3 The Local Parameter Function

Before presenting the regularity conditions ensuring the consistency of the local pseudo-maximum likelihood estimator when the number of observations  $n$  tends to infinity whereas the bandwidth  $h$  tends to zero, we derive the possible limits for  $\tilde{\theta}_{c,h}$  when  $h$  tends to zero.

**Property 5.1:**

Let us assume the conditions A1-A4 and

**A.5** The density functions  $f(y, \theta)$  and  $f_0(y)$  are positive and third order differentiable with respect to  $y$ .

**A.6** For  $h$  small and any  $c$ , the following integrals exist:  $\int K(u) \log f(c + uh; \theta) f_0(c + uh) du$ ,  $\int K(u) f_0(c + uh) du$ ,  $\int K(u) f(c + uh; \theta) du$ , and are twice differentiable under the integral sign with respect to  $h$ .

i) When  $h$  tends to zero, the objective function :

$$A_h(\theta) = E_0 \left[ \frac{1}{h} K \left( \frac{Y - c}{h} \right) \log f(Y; \theta) \right] - E_0 \left[ \frac{1}{h} K \left( \frac{Y - c}{h} \right) \right] \log \int \frac{1}{h} K \left( \frac{y - c}{h} \right) f(y; \theta) dy$$

is equivalent to :

$$A_h(\theta) = \frac{\eta^2 h^2}{2} \left[ 2 \frac{\partial \log f(c; \theta)}{\partial y} \frac{\partial \log f_0(c)}{\partial y} - \left[ \frac{\partial \log f(c; \theta)}{\partial y} \right]^2 \right] + o(h^2).$$

ii) The local parameter function (l.p.f.)  $\tilde{\theta}(c; f_0) = \lim_{h \rightarrow 0} \tilde{\theta}_{c,h}$  is the solution of the equation:

$$\frac{\partial \log f(c; \tilde{\theta}(c; f_0))}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}.$$

**Proof:**

i) is shown in Appendix 1.

ii) The objective function is a locally quadratic function with respect to

$$\frac{\partial \log f(c; \theta)}{\partial y},$$

which implies the first order condition:

$$\frac{\partial \log f(c; \theta)}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}.$$

Q.E.D.

We deduce from the previous property the following corollary:

*Corollary 5.1*

The local parameter function characterizes the distribution.

**Proof:**

Indeed, if  $\tilde{\theta}(\cdot; f_0)$  is known, we also know the log-derivative of the density function, since

$$\frac{\partial \log f_0(c)}{\partial y} = \frac{\partial \log f(c; \tilde{\theta}(c, f_0))}{\partial y}, \quad \forall c,$$

and by integration we find the density function  $f_0$  itself.

Q.E.D.

*Corollary 5.2*

The local parameter function is constant if and only if the distribution belongs to the pseudo-family  $\mathcal{F}$ .

**Proof:**

If  $f_0(y) = f(y; \theta_0)$ , we see immediately that  $\tilde{\theta}(c, f_0) = \theta_0$  is constant. The converse part results from Corollary 5.1.

Q.E.D.

The Corollary 5.1 may be used to derive the functional estimators of the log-derivative of the density function. Indeed if  $\tilde{\theta}_n(c)$  is a consistent functional estimator of  $\tilde{\theta}(c, f_0)$ , then

$$\frac{\partial \log f(c; \tilde{\theta}_n(c))}{\partial y}$$

is a consistent functional estimator of  $\frac{\partial \log f_0(c)}{\partial y}$ , whereas  $\frac{\partial \log f(c; \theta)}{\partial y}$  is continuous with respect to  $\theta$ . Also, the Corollary 5.2 may be used for testing the goodness of fit in the family  $\mathcal{F}$ .

*Example 5.1:*

For the gaussian family  $\mathcal{F} = \{f(y; \theta) = \phi(y - \theta), \theta \in R\}$ , we have:

$$\frac{\partial \log f(c; \theta)}{\partial y} = \frac{\partial \log \phi(c - \theta)}{\partial y} = \theta - c.$$

We deduce that:

$$\tilde{\theta}(c; f_0) = c + \frac{\partial \log f_0(c)}{\partial y}.$$

Q.E.D.

## 5.4 Local Parameter Function and Tail Index

In this section we show that the local parameter function can be used to find the tail index  $\gamma > 0$  of the true unknown distribution  $f_0$ . From the Karamata's theorem [Karamata (1962), Resnick (1987), Corollary 1.12], the true distribution is such that :  $1 - F_0(y) = y^{-\gamma} L(y)$ ,  $y > 1$ , with  $\gamma > 0$ , if and only if :

$$1 - F_0(y) = c(y) \exp - \int_1^y \frac{\gamma(t)}{t} dt, \quad (5.3)$$

where the  $c$  and  $\gamma$  functions satisfy:

$$\begin{aligned} \lim_{y \rightarrow \infty} c(y) &= c > 0, \\ \lim_{y \rightarrow \infty} \gamma(y) &= \gamma > 0. \end{aligned} \quad (5.4)$$

Let us assume that  $c$  and  $\gamma$  are twice differentiable and satisfy the additional limiting conditions:

$$\lim_{y \rightarrow \infty} y c'(y) = \lim_{y \rightarrow \infty} y^2 c''(y) = \lim_{y \rightarrow \infty} y \gamma'(y) = 0. \quad (5.5)$$

We get from (5.3):

$$-y \frac{\partial \log f_0(y)}{\partial y} = \gamma(y) + 1 + \frac{y c'(y) \gamma(y) + c(y) y \gamma'(y) - y c'(y) - y^2 c''(y)}{c(y) \gamma(y) - y c'(y)},$$

which implies:

$$\lim_{y \rightarrow \infty} -y \frac{\partial \log f_0(y)}{\partial y} = \gamma + 1. \quad (5.6)$$

Therefore the tail index  $\gamma$  may be directly derived from the limiting behaviour of  $-y \frac{\partial \log f_0(y)}{\partial y}$ .

Let us now consider the asymptotic behaviour of the local parameter function, which clearly depends on the tail behaviour of the pseudo-family  $f(y; \theta)$ .

i) If any distribution  $f(y; \theta)$  admits a tail index  $\gamma(\theta)$  in a one to one continuous relationship with  $\theta$ , we deduce from the equation defining  $\tilde{\theta}(c, f_0)$  that:

$$-c \frac{\partial f[c; \tilde{\theta}(c, f_0)]}{\partial y} = -c \frac{\partial \log f_0(c)}{\partial y},$$

and by taking the limit when  $c$  tends to infinity we find that:

$$\lim_{c \rightarrow \infty} \gamma[\tilde{\theta}(c, f_0)] + 1 = \gamma[\lim_{c \rightarrow \infty} \tilde{\theta}(c, f_0)] + 1 = \gamma + 1.$$

Therefore the local parameter function tends to a limit such that the tail index of the estimated pseudo-distribution coincides with the index of the true distribution.

ii) For a family of distributions with thin tails, the term  $\tilde{\theta}(c, f_0)$  may diverge to infinity when  $c$  tends to infinity. The examples below illustrate this feature.

*Example 5.2:* If the pseudo-family is the gaussian family indexed by the mean, we get:

$$-y \frac{\partial \log f(y; \theta)}{\partial y} = y(y - \theta),$$

and by Property 5.1:

$$\tilde{\theta}(y; f_0) = y - \frac{1}{y} y \frac{\partial \log f_0(y)}{\partial y} \approx y,$$

for large  $y$ .

## 5.5 Asymptotic Properties of the Local Pseudo-Maximum Likelihood Estimator

The asymptotic properties of the local P.M.L. estimator of  $\theta$  are derived along the following lines. We first find the asymptotically equivalent formula of the objective function and of the estimator, which only depend on a limited number of kernel estimators. Then, we deduce the properties of the local P.M.L. estimator from the properties of these basic kernel estimators. We only detail the additional assumptions which are necessary for the asymptotic equivalence to hold, since the set

of assumptions for the existence and asymptotic normality of the basic kernel estimators are quite standard [see Parakasa-Rao (1983), Bosq- Lecoutre(1987), Hardle (1990)].

**Property 5.2**

The local pseudo-maximum likelihood estimator  $\tilde{\theta}_n(c)$  exists and is a strongly consistent estimator of the local parameter function  $\tilde{\theta}(c; f_0)$  under A.1 - A.6 and the following additional assumptions:

**A.7** : the parameter set  $\Theta$  is an open set;

**A.8** : there exists a unique solution in  $\theta$  of the equality:

$$\frac{\partial \log f(c; \theta)}{\partial y} = \frac{\partial \log f_0(c)}{\partial y};$$

**A.9** : the following kernel estimators are strongly consistent:

$$(i) \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \xrightarrow{a.s.} f_0(c);$$

$$(ii) \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left(\frac{y_i - c}{h}\right)^2 \xrightarrow{a.s.} \eta^2 f_0(c);$$

$$(iii) \frac{1}{nh} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left(\frac{y_i - c}{h}\right) \xrightarrow{a.s.} \eta^2 f_0(c) \frac{\partial \log f_0(c)}{\partial y};$$

**A.10** In any neighbourhood of  $\theta$ , the third order derivative  $[\partial^3 \log f(y; \theta)]/\partial y^3$  is dominated by a function  $a(y)$  such that  $y^3 a(y)$  is integrable.

**Proof:**

See Appendix 2(i).

**Property 5.3**

Under assumptions A.1 - A.10 the local pseudo-maximum likelihood estimator is asymptotically equivalent to the solution  $\tilde{\theta}_n(c)$  of the equation:

$$\frac{\partial \log f[c; \tilde{\theta}_n(c)]}{\partial y} = \frac{1}{\eta^2} \frac{1}{h^2} (\tilde{m}_n(c) - c),$$

where:

$$\tilde{m}_n(c) = \frac{\sum_{i=1}^n K\left(\frac{y_i - c}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{y_i - c}{h}\right)},$$

is the Nadaraya - Watson estimator of  $m(c) = E(Y|Y = c) = c$  based on the kernel  $K$ .

**Proof:**

See Appendix 2(ii).

Therefore the asymptotic distribution of  $\tilde{\theta}_n(c)$  may be deduced from the standard properties of  $\tilde{m}_n(c) - c$ . We only consider below the pointwise convergence in distribution [A functional theorem for  $\tilde{\theta}_n(\cdot)$  may only be derived if there exists a functional theorem for  $\tilde{m}_n(c)$  at second order].

Under standard regularity conditions [Prakasa-Rao (1983), Hardle (1990), Bosq, Lecoutre (1987)] the numerator and denominator of  $1/h^2 (\tilde{m}_n(c) - c)$  have the following asymptotic properties:

**A.11** : If  $n \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $nh^3 \rightarrow \infty$ ,  $nh^5 \rightarrow 0$ , we have the limiting distribution:

$$\begin{pmatrix} \sqrt{nh^3} \left[ \frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{y_i - c}{h}\right) (y_i - c) - \eta^2 \frac{\partial f_0(c)}{\partial y} \right] \\ \sqrt{nh} \left[ \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y_i - c}{h}\right) - f_0(c) \right] \end{pmatrix} \xrightarrow{d} N \left( 0, f_0(c) \begin{bmatrix} \int u^2 K^2(u) du & \int u K^2(u) du \\ \int u K^2(u) du & \int K^2(u) du \end{bmatrix} \right).$$

The formulas of the first and second order asymptotic moments are easily verified [see Appendix 3]. The rate of convergence of the numerator is slower than the rate of convergence of the denominator since we study the degenerate case, where the Nadaraya-Watson estimator is applied to a regression with regressor equal to the regressand. We deduce that the asymptotic distribution of

$$\sqrt{nh^3} \left( \frac{1}{h^2} [\tilde{m}_n(c) - c] - \eta^2 \frac{\partial \log f_0(c)}{\partial y} \right),$$

coincides with the asymptotic distribution of

$$\sqrt{nh^3} \frac{1}{f_0(c)} \left( \frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{y_i - c}{h}\right) (y_i - c) - \eta^2 \frac{\partial f_0(c)}{\partial y} \right),$$

i.e.  $N \left[ 0, \frac{1}{f_0(c)} \int u^2 K^2(u) du \right]$ .

We deduce directly by the  $\delta$ -method the asymptotic distribution of the local pseudo maximum likelihood estimator and of the log derivative of the true p.d.f.

#### Property 5.4

Under assumptions A.1 - A.11 we have:

i)

$$\sqrt{nh^3} \left( \frac{\partial \log f(c; \tilde{\theta}_n(c))}{\partial y} - \frac{\partial \log f_0(c)}{\partial y} \right) \xrightarrow{d} N \left[ 0, \frac{1}{\eta^4 f_0(c)} \int u^2 K^2(u) du \right].$$

ii)

$$\sqrt{nh^3} \left( \tilde{\theta}_n(c) - \tilde{\theta}(c; f_0) \right) \xrightarrow{d} N \left[ 0, \left( \frac{\partial^2 \log f(c; \tilde{\theta}(c; f_0))}{\partial y \partial \theta} \right)^{-2} \frac{1}{\eta^4 f_0(c)} \int u^2 K^2(u) du \right].$$

*Remark 5.* : The functional estimator of the log-derivative  $\frac{\partial \log f_0(c)}{\partial y}$  may be compared to the standard one:

$$\frac{\partial \log \hat{f}_0(c)}{\partial y} = \frac{\sum_{i=1}^n \frac{1}{h} K' \left( \frac{y_i - c}{h} \right)}{\sum_{i=1}^n \frac{1}{h} K \left( \frac{y_i - c}{h} \right)}.$$

The rate of convergence is identical and the asymptotic distribution is:

$$\sqrt{nh^3} \left( \frac{\partial \log \hat{f}_0(c)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y} \right) \xrightarrow{d} N \left[ 0, \frac{1}{\eta^4 f_0(c)} \int K'(u)^2 du \right].$$

The asymptotic distributions of the two kernel estimators of the density function are identical if  $|K'(u)| = |uK(u)|$ , in particular for a gaussian kernel.

## 6 Empirical Results

We examine a series of returns for the Alcatel stock covering the period of July and August 1995. Alcatel belongs to the most frequently traded stocks on the Paris Stock Exchange (Paris Bourse). On average the shares of Alcatel are exchanged every 52 seconds. The data consist of 20502 observations recorded intradaily with an accuracy of 1 second. Figure 1 shows the dynamic pattern of the Alcatel series. Due to irregular spacing, the conventional time scale is replaced in Figure 1 by an axis ordered by indices of subsequent trades, or equivalently numbers of observations in the sample. The Alcatel returns feature a tendency for clustering and a time varying volatility, while potential deviations of the mean return from zero can visually not be distinguished. Figure 2 displays the marginal density of returns and indicates the examined tail area. The density function is centered at 4.259E-6, while the variance and standard deviation are 9.771E-7 and 0.000989, respectively. The distribution is slightly asymmetric, with the skewness coefficient -0.00246. The high kurtosis 5.329689 is due to the presence of heavy tails stretched between the extreme values of -0.00813 and 0.007255. 90 % of the probability mass is concentrated between -0.0017 and 0.001738. The interquartile range 0.000454 is 100 times smaller than the overall range 0.01538. The shape of tails, although kernel smoothed in Figure 2, suggests some local irregularities in the rate of tail decay. Slight lobes can clearly be distinguished in both tails. In the right tail we observe a higher probability of returns taking values between 0.0012 and 0.0013 compared to the probability of those of a slightly smaller size, i.e. between 0.0010 and 0.0012. In the left tail we recorded relatively more returns between -0.0014 and -0.0012 than those taking marginally higher values.

We perform the tail analysis using three pseudo-models:

- 1) a gaussian pseudo-model,
- 2) a Pareto pseudo-model,
- 3) an exponential pseudo-model applied to Box-Cox transformed data.

In each case, we apply the methods proposed in the paper. The first method consists in computing the pseudo maximum likelihood estimator over the interval  $[c_j, \infty)$  for an increasing sequence  $(c_j)$ ,  $j = 1, \dots, n$ . The second method involves maximization of the pseudo likelihood function over disjoint intervals  $[c_j, c_{j+1}]$ ,  $j = 1, \dots, n$ . These procedures yield estimators of the t.p.f. and the l.p.f., respectively. We present below the optimization criteria for the estimation of the t.p.f for the three pseudo-models considered.

1) **gaussian family** : The t.p.f. estimator consists of a sequence of solutions  $(\mu_c, \sigma_c^2)$ ,  $c = c_j$ ,  $j = 1, \dots, n$  of the following maximization problems:

$$\max_{\mu, \sigma} \sum_{i=1}^n 1_{y_i > c_j} \left( -\frac{1}{2} \log \sigma^2 - \frac{1}{2} \log 2\pi - \frac{(y_i - \mu)^2}{2\sigma^2} - \log \Phi \left( \frac{\mu - c}{\sigma} \right) \right)$$

for  $c_j$ ,  $j = 1, \dots, n$ .

To obtain a concave objective function, we introduce the change of parameters:  $a = 1/\sigma$ ,  $b = \mu/\sigma$ .

2) **Pareto family** : For  $\log Y_i \sim \gamma(1; \theta)$  the t.p.f is obtained by solving:

$$\max_{\theta} \sum_{i=1}^n 1_{y_i > c_j} [\log \theta - \theta(\log y_i - c_j)].$$

The estimator is:

$$\frac{1}{\hat{\theta}} = \frac{\sum_{i=1}^n 1_{y_i > c_j} (\log y_i - c_j)}{\sum_{i=1}^n 1_{y_i > c_j}}.$$

3) **Box - Cox transformed exponential variables** : We introduce a transformed variable  $X = (Y_i^\lambda - 1)/\lambda \sim \gamma(1; \theta)$ . The survivor function is defined by the probability:

$$S(y; \theta, \lambda) = \exp \left[ -\theta \frac{y_i^\lambda - 1}{\lambda} \right],$$

while the density of the pseudo model is:

$$f(y; \theta, \lambda) = \theta y_i^{\lambda-1} \exp \left[ -\theta \frac{y_i^\lambda - 1}{\lambda} \right].$$



To estimate the t.p.f we maximize the objective function:

$$\max_{\lambda, \theta} \sum_{i=1}^n 1_{y_i > c_j} \left[ \log \theta + (\lambda - 1) \log y_i - \theta \frac{y_i^\lambda - 1}{\lambda} + \theta \frac{c_j^\lambda - 1}{\lambda} \right].$$

The optimization in  $\theta$  can be performed analytically.

The estimation of the local parameter function involves the following maximizations:

$$\max_{\theta} \sum_{i=1}^n 1_{c_j < y_i < c_{j+1}} [\log f(y_i; \theta) - \log [S(c_j; \theta) - S(c_{j+1}; \theta)]],$$

over a sequence of non overlapping intervals  $(c_j, c_{j+1})$ , with  $f$  and  $S$  replaced by adequate expressions corresponding to the selected pseudo - family.

The right tail of Alcatel data was examined using an increasing  $c_j$  sequence of centiles of the return density. We set the first term,  $c_1$  equal to the 76th centile, (return value of 0.00023) while the last to the maximum of the data, with an increment of two centiles, i.e. 410 observations. Accordingly, the t.p.f. was first estimated on the interval between the 76th and 100th centiles ending up with the interval between the 98th and 100th centiles. The l.p.f was estimated over a sequence of intervals containing each 410 observations, proceeding with the intervals between the 76th and 78th centile, 78th and 80th, etc.

The estimation of the t.p.f. for the gaussian pseudo-family is illustrated in Figures 3.a and 3.b. As expected, the estimator of  $1/\sigma$  is close to zero indicating the existence of heavy tails. The non-smooth t.p.f. of the  $\mu/\sigma$  parameter, displayed in the top panel, is due to the presence of slight lobes in the density function. Figures 4.a and 4.b show the t.p.f and l.p.f. of  $\log(\theta)$ , the logarithm of the parameter of the Pareto pseudo-family. We observe that the estimated t.p.f does not admit an asymptote. Moreover, its smooth increase over the sequence of semi-intervals results from an aggregation of very heterogenous values estimated in disjoint intervals, displayed below. Finally, Figures 5.a-5.b correspond to the Box-Cox transformed Pareto family and report again the logarithmic values of parameters. We find that the Pareto family, underlying the definition of the Hill estimator, is strongly misspecified since  $\log \lambda$  is significantly negative.

Next, we estimated the log-derivative of the density function

$$c \rightarrow -\frac{c}{h^2} [\tilde{m}_n(c) - c] = -\frac{c}{h^2} \frac{\sum_{i=1}^n K\left(\frac{y_i - c}{h}\right) (y_i - c)}{\sum_{i=1}^n K\left(\frac{y_i - c}{h}\right)},$$

using a gaussian kernel. We show in Figure 6, the results from estimating a set of 10000 generated normally distributed variables with mean 5 and variance 1. Figure 6 clearly reveals the linear affine form of the underlying function. The log-derivative of Alcatel density is presented in Figure 7. At

the extremes, the functional estimator approaches non-zero asymptotes of almost equal absolute values, suggesting a symmetric, exponential decay rate of both density tails. The decay rate is faster than it would be for Pareto tails, but slower than the decay rate of the normal density. This can easily be shown by integrating:

$$f(y) = \int \frac{\partial \log f(y)}{\partial y} dy = \int k dy \approx \exp(ky)$$

where  $k$  denotes the constant representing the upper or lower bound of the functional estimator. Clearly, the data feature an exponential rate of tail decay, slower than a normally distributed variable.

## 7 Conclusions

We proposed a method of tail analysis based on the truncated pseudo-maximum likelihood. This approach was extended to a goodness of fit test and a kernel estimator of the log-derivative of the density function. The procedures and the asymptotic results were presented in the framework of i.i.d. variables, for which the marginal and conditional distribution coincide.

We realize that in some applications to, for example, financial data, dynamic patterns like the serial correlation and conditional heteroscedasticity need to be accommodated. Since the risk analysis is based on the conditional distributions, it appears necessary to extend the techniques proposed in this paper to a dynamic framework, i.e. to find an approximation of the true conditional distribution by means of a pseudo-family of conditional distributions. This extension, remaining a topic of further research, requires in particular an interpretation of the local parameter function for parameters such as the autocorrelations and autocorrelations of the volatilities.

## Appendix 1

We derive the expansion of the objective function :

$$A_h(\theta) = E_0 \left[ \frac{1}{h} K \left( \frac{Y-c}{h} \right) \log f(Y; \theta) \right] - E_0 \left[ \frac{1}{h} K \left( \frac{Y-c}{h} \right) \right] \log \int \frac{1}{h} K \left( \frac{y-c}{h} \right) f(y; \theta) dy,$$

when  $h$  approaches zero.

By introducing the true density function  $f_0(y)$  we obtain:

$$A_h(\theta) = \int \frac{1}{h} K \left( \frac{y-c}{h} \right) \log f(y; \theta) f_0(y) dy - \int \frac{1}{h} K \left( \frac{y-c}{h} \right) f_0(y) dy \log \int \frac{1}{h} K \left( \frac{y-c}{h} \right) f(y; \theta) dy.$$

Let us denote the ratio  $(y-c)/h = u$ , so that  $y = c + uh$ . By substituting in the equation above, the objective function can be rewritten:

$$A_h(\theta) = \int K(u) \log f(c + uh; \theta) f_0(c + uh) du - \int K(u) f_0(c + uh) du \log \int K(u) f(c + uh; \theta) du.$$

For  $h = 0$  we have:

$$\begin{aligned} A_0(\theta) &= \int K(u) \log f(c; \theta) f_0(c) du - \int K(u) f_0(c) du \log \int K(u) f(c; \theta) du \\ &= f_0(c) \log f(c; \theta) - f_0(c) \log f(c; \theta) = 0, \end{aligned}$$

by assumption A.2.

The first order Taylor expansion in the neighborhood of  $h = 0$  yields  $A_h(\theta) \approx 0$  given that  $\int uK(u)du = 0$  by Assumption A.3. The second order expansion yields :

$$\begin{aligned} A_h(\theta) \approx & \int K(u) \left[ \log f(c; \theta) + \frac{\partial \log f(c; \theta)}{\partial y} uh + \frac{\partial^2 \log f(c; \theta)}{\partial y^2} \frac{u^2 h^2}{2} \right] \\ & \left[ f_0(c) + \frac{\partial f_0(c)}{\partial y} uh + \frac{\partial^2 f_0(c)}{\partial y^2} \frac{u^2 h^2}{2} \right] du \\ & - \int K(u) \left[ f_0(c) + \frac{\partial f_0(c)}{\partial y} uh + \frac{1}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \frac{u^2 h^2}{2} \right] du \\ & \log \left[ \int K(u) \left[ f(c; \theta) + \frac{\partial f(c; \theta)}{\partial y} uh + \frac{\partial^2 f(c; \theta)}{\partial y^2} \frac{u^2 h^2}{2} \right] du \right]. \end{aligned}$$

Let us denote  $\eta^2 = \int u^2 K(u) du$ .

$$A_h(\theta) \approx f_0(c) \log f(c; \theta) + \frac{\eta^2 h^2}{2} \left[ \log f(c; \theta) \frac{\partial^2 f_0(c)}{\partial y^2} + 2 \frac{\partial \log f(c; \theta)}{\partial y} \frac{\partial f_0(c)}{\partial y} + f_0(c) \frac{\partial^2 \log f(c; \theta)}{\partial y^2} \right]$$

$$- \left[ f_0(c) + \frac{\eta^2 h^2}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \right] \log \left[ f(c; \theta) + \frac{\eta^2 h^2}{2} \frac{\partial^2 f(c; \theta)}{\partial y^2} \right].$$

Consider the second term in the last expression. Explicitly, it can be represented as :

$$\begin{aligned} & - \left[ f_0(c) + \frac{\eta^2 h^2}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \right] \left[ \log f(c; \theta) + \log \left[ 1 + \frac{\eta^2 h^2}{2} \frac{\partial^2 f(c; \theta)}{\partial y^2} / f(c; \theta) \right] \right] \\ \approx & - \left[ f_0(c) + \frac{\eta^2 h^2}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \right] \left[ \log f(c; \theta) + \frac{\eta^2 h^2}{2} \frac{\partial^2 f(c; \theta)}{\partial y^2} / f(c; \theta) \right] \\ \approx & - f_0(c) \log f(c; \theta) - \frac{\eta^2 h^2}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \log f(c; \theta) - \frac{\eta^2 h^2}{2} f_0(c) \frac{\partial^2 f(c; \theta)}{\partial y^2} / f(c; \theta). \end{aligned}$$

By collecting both terms,  $A_h(\theta)$  becomes:

$$A_h(\theta) \approx \frac{\eta^2 h^2}{2} \left[ 2 \frac{\partial \log f(c; \theta)}{\partial y} \frac{\partial f_0(c)}{\partial y} + f_0(c) \left[ \frac{\partial^2 \log f(c; \theta)}{\partial y^2} - \frac{\partial^2 f(c; \theta)}{\partial y^2} / f(c; \theta) \right] \right].$$

Note that :

$$\begin{aligned} \frac{\partial^2 \log f}{\partial y^2} &= \frac{\partial}{\partial y} \left[ \frac{\partial \log f}{\partial y} \right] = \frac{\partial}{\partial y} \left[ \frac{f'}{f} \right] \\ &= \frac{f''}{f} - \frac{f'^2}{f^2}, \end{aligned}$$

and hence,

$$A_h(\theta) \approx \frac{\eta^2 h^2}{2} \left[ 2 \frac{\partial \log f(c; \theta)}{\partial y} \frac{\partial f_0(c)}{\partial y} - f_0(c) \left[ \frac{\partial \log f(c; \theta)}{\partial y} \right]^2 \right],$$

or, equivalently

$$A_h(\theta) \approx \frac{\eta^2 h^2}{2} f_0(c) \left[ 2 \frac{\partial \log f(c; \theta)}{\partial y} \frac{\partial \log f_0(c)}{\partial y} - \left( \frac{\partial \log f(c; \theta)}{\partial y} \right)^2 \right].$$

Note also that  $A_h(\theta)$  is represented in the last equation as a quadratic function of  $\frac{\partial \log f(c; \theta)}{\partial y}$ .

Therefore the maximization with respect to  $\theta$  involves only this term and the optimizing value  $\tilde{\theta}(c; f_0)$  of  $\theta$  satisfies :

$$\frac{\partial \log f(c; \tilde{\theta}(c; f_0))}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}.$$

## Appendix 2

### (i) The Consistency

Let us consider the normalized objective function:

$$\tilde{A}_{n,h}(\theta) = \frac{1}{nh^2} \left[ \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \log f(y_i; \theta) - \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \log \int \frac{1}{h} K\left(\frac{y - c}{h}\right) f(y; \theta) dy \right].$$

It can be written as:

$$\begin{aligned} \tilde{A}_{n,h}(\theta) &= \frac{1}{nh^2} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left[ \log f(c; \theta) + \frac{\partial \log f(c; \theta)}{\partial y} (y_i - c) + \frac{1}{2} \frac{\partial^2 \log f(c; \theta)}{\partial y^2} (y_i - c)^2 \right. \\ &\quad \left. + R_1(y_i; \theta) (y_i - c)^3 \right] \\ &\quad - \frac{1}{nh^2} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left[ \log f(c; \theta) + \frac{h^2 \eta^2}{2} \frac{1}{f(c; \theta)} \frac{\partial^2 f(c; \theta)}{\partial y^2} + h^3 R_2(\theta, h) \right], \end{aligned}$$

where  $R_1(y_i; \theta), R_2(\theta)$  are the residual terms in the expansion. We deduce:

$$\begin{aligned} \tilde{A}_{n,h}(\theta) &= \frac{1}{nh^2} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) (y_i - c) \frac{\partial \log f(c; \theta)}{\partial y} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left(\frac{y_i - c}{h}\right)^2 \frac{1}{2} \frac{\partial^2 \log f(c; \theta)}{\partial y^2} \\ &\quad - \frac{\eta^2}{2} \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \frac{\frac{\partial^2 f}{\partial y^2}(c; \theta)}{f(c; \theta)} + \text{residual terms.} \end{aligned}$$

Under the assumptions of Property 5.2, the residual terms tend almost surely to zero, uniformly on  $\Theta$ , whereas the main terms tend almost surely uniformly on  $\Theta$  to:

$$A_\infty(\theta) = \frac{\eta^2}{2} f_0(c) \left[ 2 \frac{\partial \log f_0(c)}{\partial y} \frac{\partial \log f(c; \theta)}{\partial y} - \left( \frac{\partial \log f(c; \theta)}{\partial y} \right)^2 \right].$$

Then, by the identifiability condition, we conclude that the estimator  $\tilde{\theta}_n(c)$  exists and is strongly consistent of  $\theta(c; f_0)$ .

### (ii) Asymptotic Equivalence

The main part of the objective function may also be written as:

$$\begin{aligned} \tilde{A}_{n,h}(\theta) &\approx \frac{1}{nh^2} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) (y_i - c) \frac{\partial \log f(c; \theta)}{\partial y} \\ &\quad - \frac{\eta^2}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) \left[ \frac{\partial \log f(c; \theta)}{\partial y} \right]^2. \end{aligned}$$

We deduce that the local parameter function can be asymptotically replaced by the solution  $\tilde{\theta}_n(c)$  of:

$$\frac{\partial \log f(c; \tilde{\theta}_n(c))}{\partial y} = \frac{1}{\eta^2} \frac{\frac{1}{nh^2} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right) (y_i - c)}{\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y_i - c}{h}\right)}.$$

## Appendix 3

### The first and second order asymptotic moments.

We restrict the analysis to the numerator term  $\frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{Y_i - c}{h}\right) (Y_i - c)$ , which implies the non standard rate of convergence.

#### (1) First Order Moment

We get:

$$\begin{aligned}
 E \left[ \frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] &= \frac{1}{h^3} E \left[ K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] \\
 &= \frac{1}{h^3} \int K\left(\frac{y - c}{h}\right) (y - c) f_0(y) dy \\
 &= \frac{1}{h} \int K(u) u f_0(c + uh) du \\
 &= \frac{1}{h} \int K(u) u \left[ f_0(c) + uh \frac{\partial f_0(c)}{\partial y} + \frac{u^2 h^2}{2} \frac{\partial^2 f_0(c)}{\partial y^2} + o(h^2) \right] du \\
 &= \eta^2 \frac{\partial f_0(c)}{\partial y} + \frac{h}{2} \frac{\partial^2 f_0(c)}{\partial y^2} \int u^3 K(u) du + o(h).
 \end{aligned}$$

#### (2) Asymptotic Variance

We have:

$$\begin{aligned}
 V \left[ \frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] &= \frac{1}{nh^6} V \left[ K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] \\
 &= \frac{1}{nh^6} \left\{ E \left[ K^2\left(\frac{Y_i - c}{h}\right) (Y_i - c)^2 \right] - \left( E \left[ K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] \right)^2 \right\} \\
 &= \frac{1}{nh^3} \left[ \int K^2(u) u^2 f_0(c + uh) du - h^3 \eta^4 \left( \frac{\partial f_0(c)}{\partial y} \right)^2 \right] \\
 &= \frac{1}{nh^3} f_0(c) \int u^2 K^2(u) du + o\left(\frac{1}{nh^3}\right),
 \end{aligned}$$

which provides the rate of convergence  $(nh^3)^{-\frac{1}{2}}$  of the standard error. Moreover the second term of the bias will be negligible if  $h(nh^3)^{\frac{1}{2}} \rightarrow 0$ , or  $nh^5 \rightarrow 0$ .

#### (3) Asymptotic Covariance

Finally we have also to consider:

$$\begin{aligned}
 Cov \left[ \frac{1}{nh^3} \sum_{i=1}^n K\left(\frac{Y_i - c}{h}\right) (Y_i - c) \right] &, \quad \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Y_i - c}{h}\right) \\
 &= \frac{1}{nh^4} Cov \left[ K\left(\frac{Y_i - c}{h}\right) (Y_i - c), K\left(\frac{Y_i - c}{h}\right) \right]
 \end{aligned}$$



$$\begin{aligned}
&= \frac{1}{nh^4} \{ E \left[ K^2 \left( \frac{Y_i - c}{h} \right) (Y_i - c) \right] \right. \\
&\quad \left. - E \left[ K \left( \frac{Y_i - c}{h} \right) (Y_i - c) \right] E K \left( \frac{Y_i - c}{h} \right) \right\} \\
&= \frac{1}{nh^4} \{ h^2 \int K^2(u) u f_0(c + uh) du + O(h^4) \} \\
&= \frac{1}{nh^2} f_0(c) \int u K^2(u) du + o \left( \frac{1}{nh^2} \right).
\end{aligned}$$

FIG 1 : Returns, Alcatel

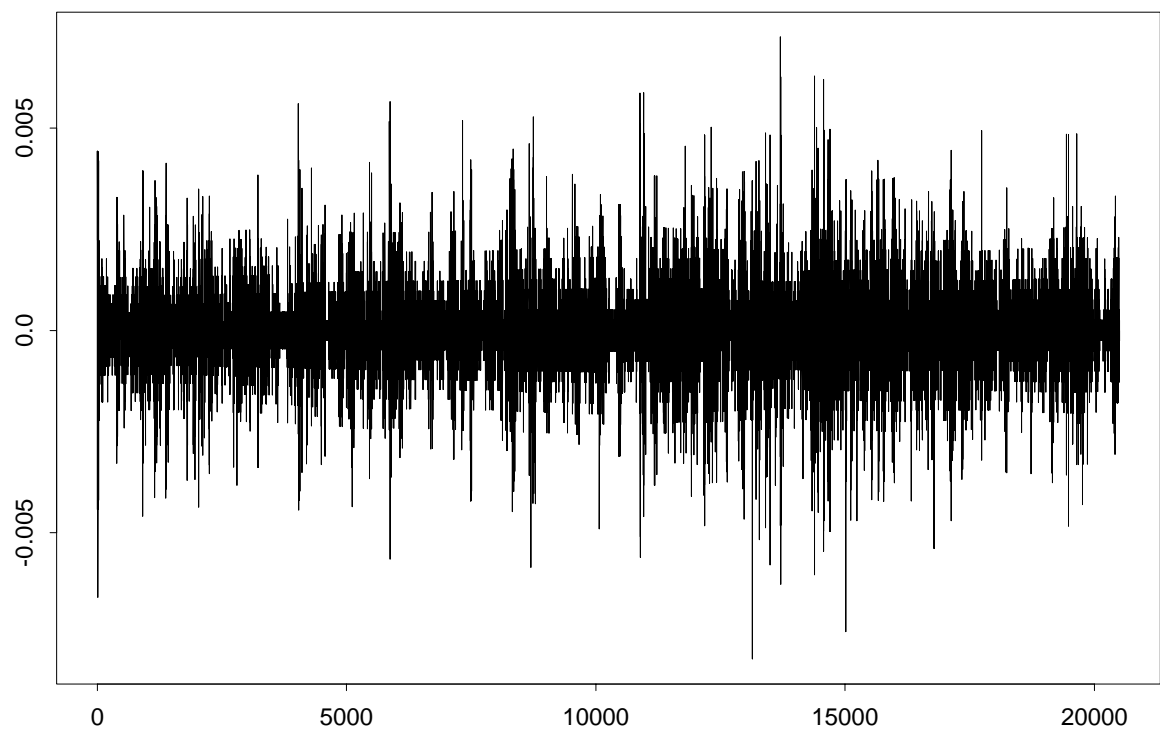


FIG 2 : Density of Returns, Alcatel

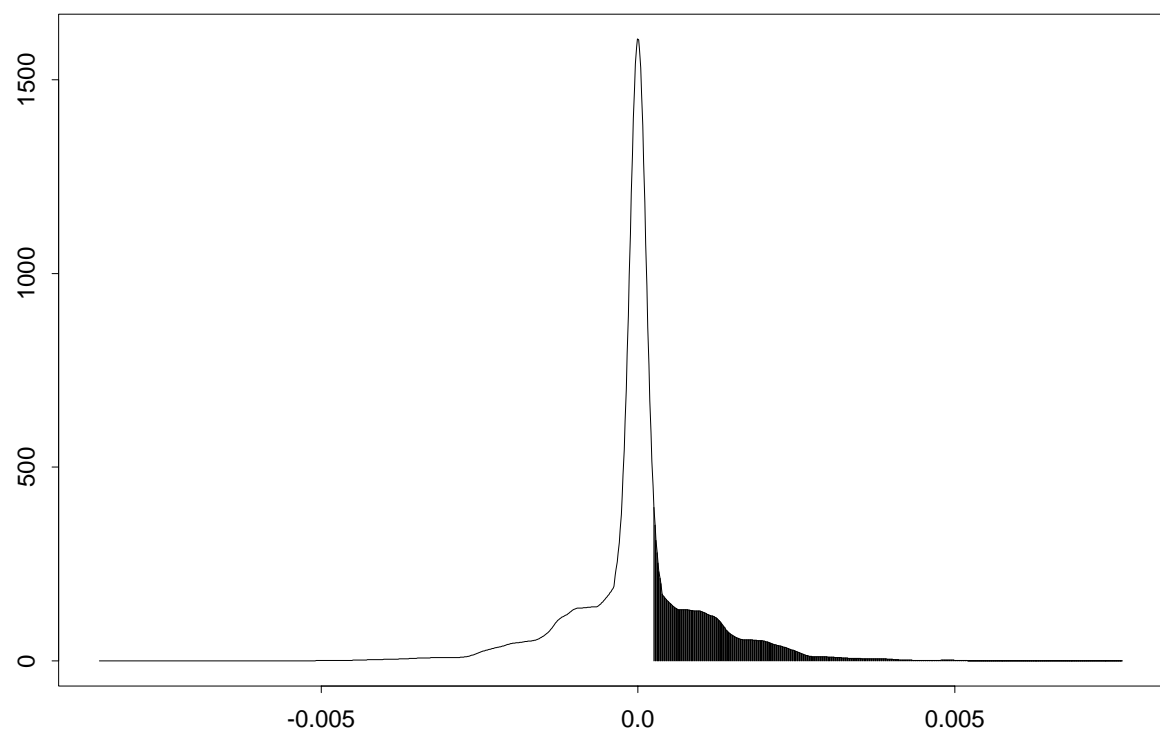


FIG 3a: Gaussian \_\_\_\_\_ mean/s.dev.

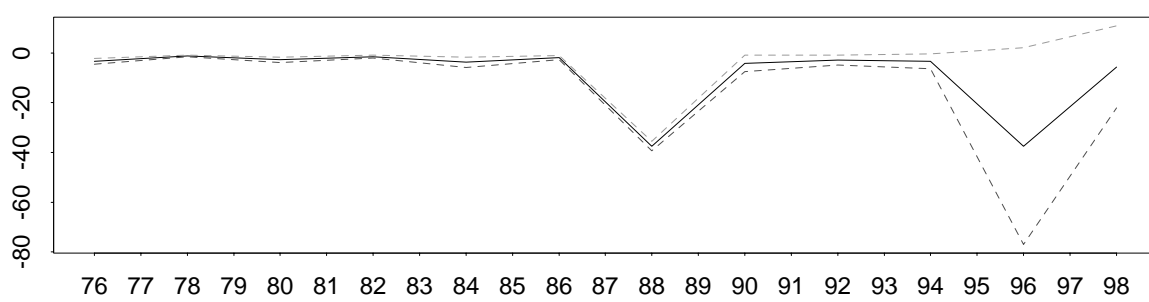


FIG 3b: Gaussian \_\_\_\_\_ 1/s.dev.

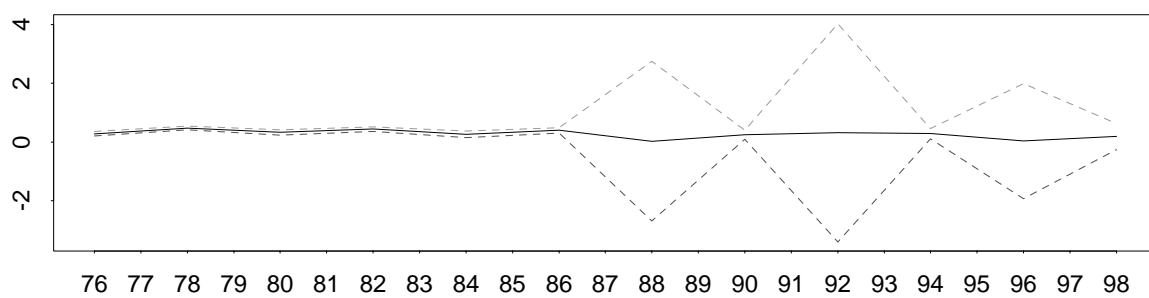


FIG 4a: Pareto \_\_\_\_\_ theta

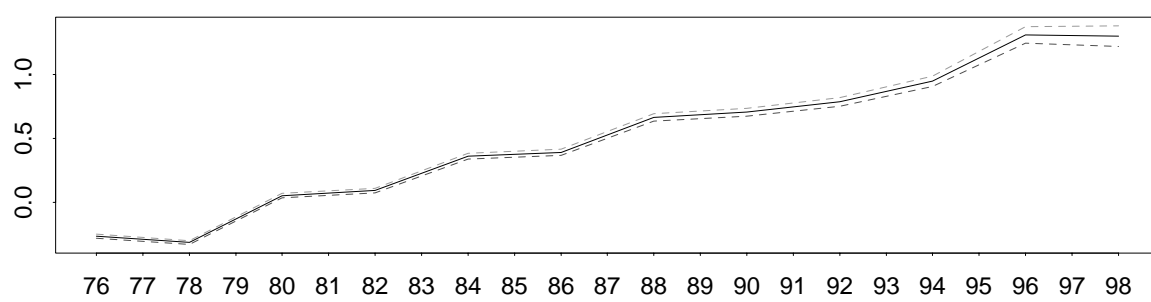


FIG 4b: Pareto \_\_\_\_\_ theta

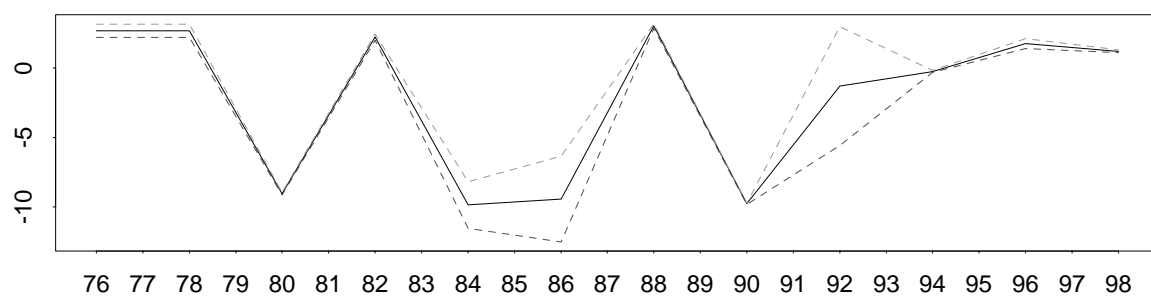


FIG 5a: Box-Cox \_\_\_\_\_ lambda

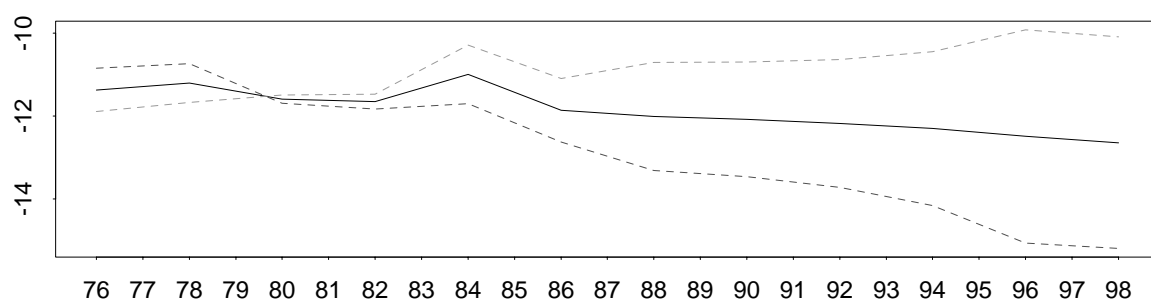


FIG 5b: Box-Cox \_\_\_\_\_ theta

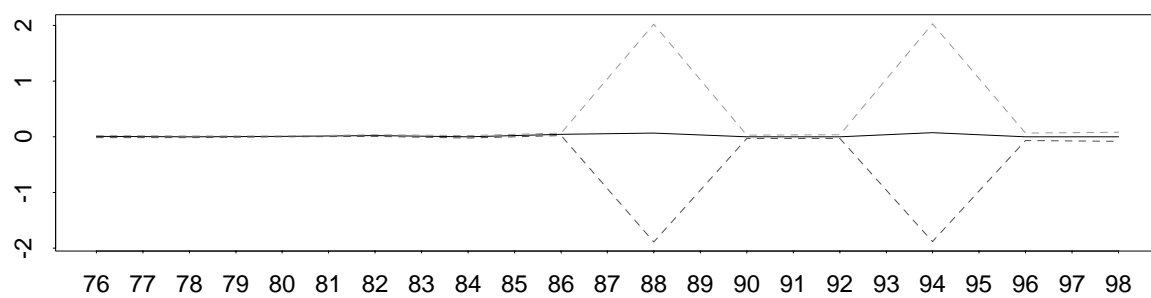


FIG 5c: Box-Cox \_\_\_\_\_ lambda

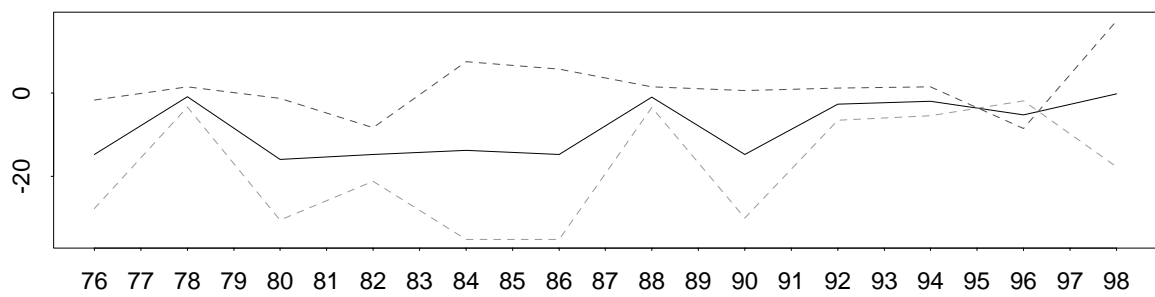


FIG 5d: Box-Cox \_\_\_\_\_ theta

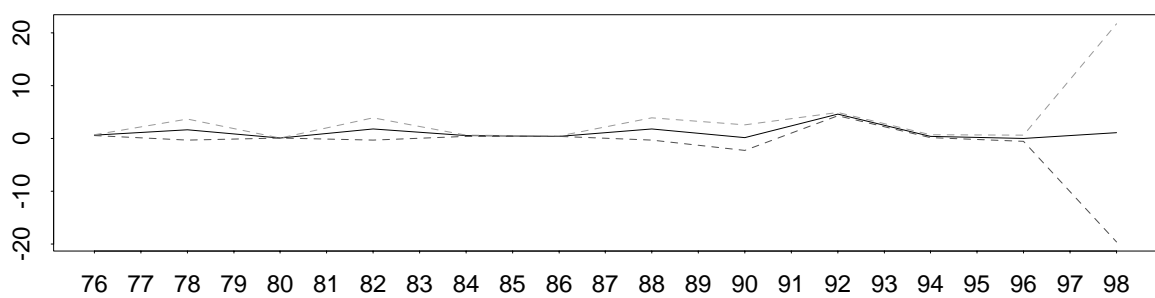


FIG 6 : Gaussian Log Derivative

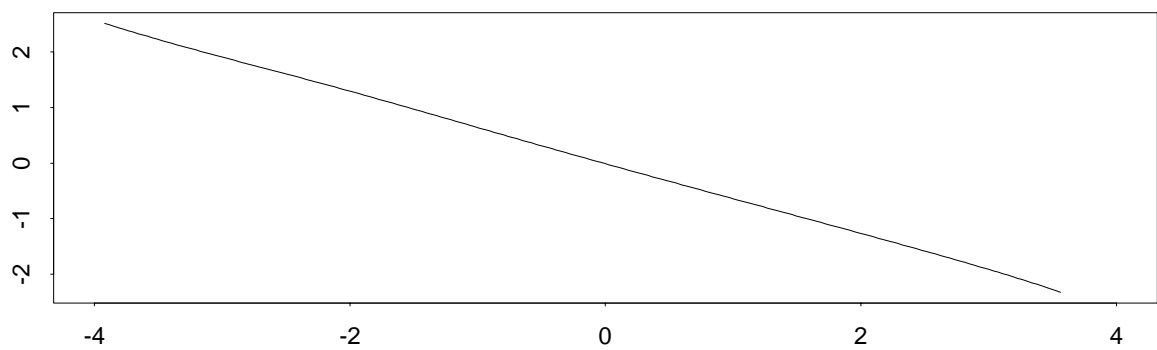
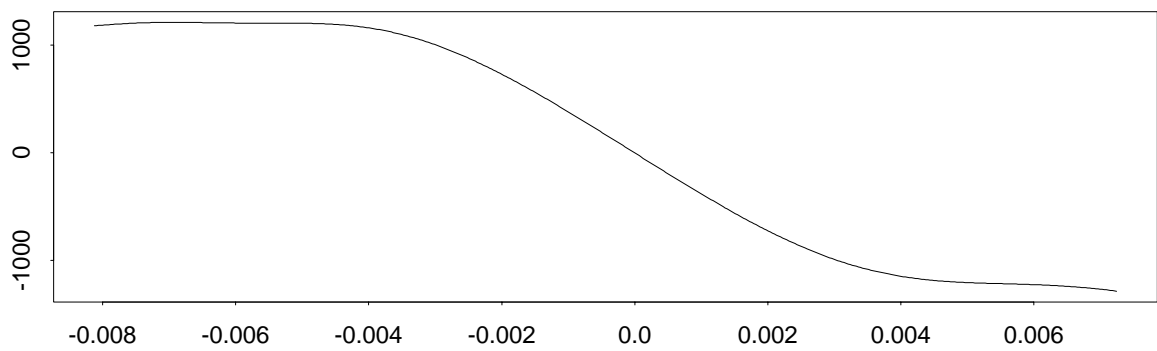


FIG 7 : Alcatel Log Derivative





## References

- [1] Berlant, J., Vynckier, P., and J. Teugels (1996) : "Tail Index Estimation, Pareto Quantile Plots and Regression Diagnostics", *Journal of the American Statistical Association*, **91**, 1659 - 1667.
- [2] Bosq, D. and J.P. Lecoutre (1987) : "Théorie de l'estimation fonctionnelle", Economica, Paris.
- [3] Clark, P. (1973) : "A Subordinated Stochastic Processes Model with Finite Variance for Speculative Prices", *Econometrica*, **41**, 135 - 156.
- [4] Cleveland, W. (1979) : "Robust Locally Weighted Regression and Smoothing Scatterplots", *JASA*, **74**, 823 - 836.
- [5] Dacarogna, M., Muller, U., Pictet, O., and C. de Vries (1994) : "The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets", Tinbergen Institute, Discussion Paper, **95-70**.
- [6] Danielsson, J. and C. de Vries (1997) : "Tail Index and Quantile Estimation with Very High Frequency Data", *Journal of Empirical Finance*, **4**, 241 - 257.
- [7] De Haan, L. (1970) : "On Regular Variation and its Application to the Weak Convergence of Sample Extremes", Math. Centre Tract, , **2**, Amsterdam.
- [8] De Haan, L. (1991) : " Extreme Value Statistics", Lecture Notes, Econometric Institute, Erasmus University, Amsterdam.
- [9] De Haan, L., Janses, D., Koedjik, K., and C. de Vries (1994) : "Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risk", in : *Extreme Value Theory and Applications*, ed. Galambos, Kluwer, 471 - 487.
- [10] Engle, R. (1982) : "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica*, **50**, 987 - 1008.
- [11] Feller, W. (1971) : "An Introduction to Probability Theory and its Applications", Vol 2, 2nd Ed., Wiley, New York.
- [12] Gouriéroux, C. and A. Monfort (1997) : " Statistics and Econometric Models", Volume 2, Cambridge University Press.
- [13] Gouriéroux, C., Monfort, A. and C. Tenreiro (1994) : "Kernel M-estimators: Non Parametric Diagnostics for Structural Models", Discussion Paper, CEPREMAP **9405**.
- [14] Hardle, W. (1990) : "Applied Nonparametric Regression", Cambridge University Press.

- [15] Hardle, W., and O., Linton (1993): "Applied Nonparametric Methods", University of Berlin, Discussion Paper 9312.
- [16] Harris, L. (1989) : "S&P 500 Cash Stock Price Volatilities", *Journal of Finance*, **44**, 1155 - 1175.
- [17] Hill, B. (1975) : "A Simple Approach to Inference About the Tail of a Distribution ", *The Annals of Statistics*, **3**, 1163 - 1174.
- [18] Jansen, D., and C. de Vries (1991) : "On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective", *Review of Economic and Statistics*, **73**, 18 - 24.
- [19] Karamata, J. (1962) : "Some Theorems Concerning Slowly Varying Functions", Mathematics Research Center, Tech. Report 369, University of Wisconsin, Madison.
- [20] Kearus, P., and A. Pagan (1997) : "Estimating the Density Tail Index for Financial Time Series", *The Review of Economics and Statistics*, **79**, 171 - 175.
- [21] Koedijk, K., Schafgans, M., and C. De Vries (1990) : " The Tail Index of Exchange Rate Returns", *Journal of International Economics*, **29**, 93 - 108.
- [22] Mandelbrot, B. (1963) : "The Variation of Certain Speculative Prices", *Journal of Business*, **36**, 394 - 419.
- [23] Mason, D. (1982) : "Laws of Large Numbers for Sums of Extreme Values", *Annals of Probability*, **10**, 754 - 764.
- [24] Nadaraya, E. (1964) : "On Estimating Regression", *Theory Prob. Appl.*, **10**, 186 - 190.
- [25] Nelson, O. (1990) : "Stationarity and Persistence in the GARCH (1,1) Model", *Econometric Theory*, **6**, 318 - 334.
- [26] Pictet, O., Dacorogna, M., and U. Muller (1996) : "Hill, Bootstrap and Jackknife Estimators for Heavy Tails", Olsen & Associates, Working Paper.
- [27] Prakasa Rao, B. (1983): "Nonparametric Functional Estimation", Academic Press.
- [28] Resnick, S. (1987) : "Extreme Values, Regular Variation and Point Processes", Springer Verlag, New York.
- [29] Resnick, S. (1997) : " Heavy Tail Modeling and Teletraffic Data ", *The Annals of Statistics*, **25**, 1805 - 1869.

- [30] Resnick, S., and G. Samorodnitsky (1997) : "Performance Decay in a Single Server Exponential Queuing Model with Long Range Dependence", *Operational Research* , forthcoming.
- [31] Rosenblatt, M. (1969) : " Conditional Probability Density and Regression Estimators", in: *Multivariate Analysis II*, 25 - 31, New York, Academic Press,
- [32] Silverman, B. (1986): "Density Estimation for Statistics and Data Analysis", London, Chapman and Hall.
- [33] Watson, G. (1964) : "Smooth Regression Analysis", *Sankhya*, Series A, **26**, 359 - 372.